Journal of Saudi Chemical Society (2014) **18**, 35–45

King Saud University

Journal of Saudi Chemical Society

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Insight into the structural requirement of substituted quinazolinone biphenyl acylsulfonamides derivatives as Angiotensin II AT_1 receptor antagonist: 2D and 3D QSAR approach

Mukesh C. Sharma *, D.V. Kohli

Drug Research Laboratory, Department of Pharmaceutical Sciences, Dr. H.S. Gour University, Sagar 470 003, Madhya Pradesh, India

Received 25 April 2011; accepted 22 May 2011
Available online 30 May 2011

KEYWORDS

Ang II;
2D QSAR;
3D kNN-MFA;
Acylsulfonamides;
AT_1;
Antihypertensive;
VLife MDS

Abstract A series of 19 molecules substituted quinazolinone biphenyl acylsulfonamides derivatives displaying variable inhibition of Angiotensin II receptor AT_1 activity were selected to develop models for establishing 2D and 3D QSAR. The compounds in the selected series were characterized by spatial, molecular and electro topological descriptors using QSAR module of Molecular Design Suite (VLife MDS™ 3.5). The best 2D QSAR model was selected, having correlation coefficient r^2 (0.8056) and cross validated squared correlation coefficient q^2 (0.6742) with external predictive ability of pred_r^2 0.7583 coefficient of correlation of predicted data set (pred_r^2se) 0.2165. The results obtained from QSAR studies could be used in designing better Ang II activity among the congeners in future. The optimum QSAR model showed that the parameters SsssCHE index, SddCE-index, T_2_Cl_4, and SssNHE-index contributed in the model. 3D QSAR analysis by kNN-molecular field analysis approach developed based on principles of the *k*-nearest neighbor method combined with Genetic algorithms, stepwise forward variable selection approach; a leave-one-out cross-validated correlation coefficient (q^2) of 0.6516 and a non-cross-validated correlation coefficient (r^2) of 0.8316 and pred_r^2 0.6954 were obtained. Contour maps using this

* Corresponding author. Tel.: +91 9826372944.
E-mail address: mukeshcsharma@yahoo.com (M.C. Sharma).

Peer review under responsibility of King Saud University.
doi:10.1016/j.jscs.2011.05.011

Production and hosting by Elsevier

approach showed that steric, electrostatic, and hydrophobic field effects dominantly determine binding affinities. The information rendered by 3D QSAR models may lead to a better understanding of structural requirements of Angiotensin II receptor and can help in the design of novel potent antihypertensive molecules.

1. Introduction

The renin-angiotensin system (RAS) is known to play a pivotal role in the regulation of fluid, electrolyte balance, and blood pressure, and is a modulator of cellular growth and proliferation. The peptide hormone Angiotensin II (Ang II) is known to play an important role in the regulation of blood pressure and salt-water homeostasis (Dzau and Pratt, 1986). Two distinct subtypes of the Ang II receptor labeled as AT_1 and AT_2 are known. Numerous studies in the past few years have shown that the major biochemical and functional responses of Ang II are mediated by activation of the AT_1 receptor. Ang II affects most of the biological functions by activating selective membrane bound receptors. Two distinct subtypes of Ang II receptors [type 1 (AT_1) and type 2 (AT_2)] have been identified, and both belong to the G protein-coupled receptors super family (GPCRs) (Tuccinardi et al., 2006). The blockade of the Ang II synthesis by converting enzyme inhibitors (CEI) has become a well-established therapeutic application in the treatment of hypertension and heart failure (Ferrario, 1990). ACE inhibitors (ACEIs) are the first class of marketing drugs targeting to the RAS system, yet all ACEIs can cause hypotension, hyperkalemia, and dry cough. Afterward, research efforts on Angiotensin receptor blocks which can replace ACE inhibitors as anti-hypertension with less adverse effects have led to the discovery of Angiotensin II antagonists (Ito, 2005). In recent decades, several selective antagonists have been developed, which are used to treat both hypertension and damage associated with diseases such as arthrosclerosis and diabetes (Ellingboe et al., 1994; Yanagisawa et al., 1996). Eight compounds have been marketed, including losartan, valsartan, candesartan, irbesartan, olmesartan and telmisartan (Burnier and Brunner, 1997). Computational chemistry, prediction of biological activity based quantitative structure–activity relationship (QSAR) substantially increases the potential of work, avoiding time, and resource consuming experiments (Dlaz and Prado, 2008). A number of quantitative structure–activity relationship (QSAR) studies related to design of Angiotensin II receptor antagonists drugs have also been reported (Belvisi et al., 1996; Datar et al., 2004; Kurup et al., 2001; Sharma et al., 2009, 2011; Sharma and Kohli, 2011a,b; Yan et al., 2007). Several 3D-QSAR techniques such as comparative molecular field analysis (COMFA), comparative molecular similarity analysis (COMSIA), and *k*-nearest neighbor (kNN) (Cramer et al., 1988; Klebe et al., 1994; Ajmani et al., 2006) are being used in modern QSAR research. The purpose of the present study is to investigate the physico-chemical parameters responsible for the Angiotensin II receptor activity of substituted quinazolinone biphenyl acylsulfonamides derivatives, explore the correlation between them and obtain more information for designing novel substituted quinazolinone biphenyl acylsulfonamides derivatives with potent antihypertensive activity. In the present investigation, three widely used techniques, viz. stepwise forward variable selection method and Genetic algorithm have been applied for descriptor optimization and partial least square analysis has been applied for two and three-dimensional QSAR models development. The generated models provide insight into the influence of various interactive fields on the activity and, thus, can help in designing and forecasting the inhibition activity of novel antihypertensive molecules.

2. Methodology

All molecular modeling studies (2D and 3D) were performed using the Molecular Design Suite (VLife MDS software package, version 3.5; from VLife Sciences, Pune, India), on a Compaq PC with a Pentium IV processor and a Windows XP operating system. Structures were sketched using the 2D draw application and converted to 3D structures.

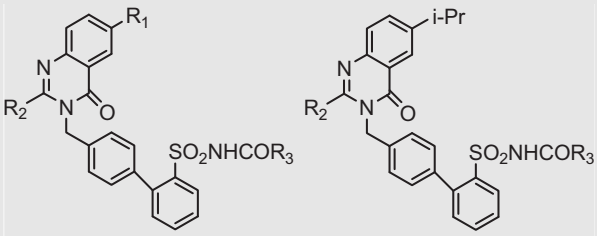
2.1. Biological data

The Angiotensin II antagonist activity data of substituted quinazolinone biphenyl acylsulfonamides derivatives were taken from the reported work (Chakravarty et al., 1994). The total set of inhibitors was divided into a training set (14 compounds) for generating 2D and 3D QSAR models and a test set (five compounds) for validating the quality of the models. Selection of the training set and test set molecules was done on the basis of structural diversity and a wide range of activity such that the test-set molecules represent a range of biological activity similar to that of the training set; thus, the test set is truly representative of the training set. The biological activity values [IC_{50} (nM)] reported in nanomolar units were converted to their molar units and then further to negative logarithmic scale and subsequently used as the dependent variable for the QSAR analysis. The $-\log$ values of IC_{50} (pIC_{50}) along with the structure of the compounds in the series are listed in Table 1.

2.2. Data set and molecular modeling for 2D QSAR

In 2D QSAR analysis, significant methods partial least square analysis, were applied to generate three models: Models I, II, and III, respectively. The 2D structures were converted to 3D structures by sending them to MDS. Each compound was energy minimized and batch optimized by using Merck Molecular Force Field (Halgren, 1996) force field and charges followed by Austin Model-1 (Gasteiger and Marsili, 1980) Hamiltonian method was available in MOPAC module with the convergence criterion 0.001 kcal/mol Å fixing Root Mean Square Gradients (RMS) to 0.001 kcal/mol Å. 2D descriptors (physicochemical and alignment independent) were calculated for the optimized compounds on QSAR plus work sheet. The invariable descriptors (the descriptors that are constant for all the molecules) were removed, as they do not contribute

Table 1 Structure and activities of quinazolinone biphenyl acylsulfonamides



Comp	R_1	R_2	R_3	AT_1IC_{50}[a] (nM)	pIC_{50}[b]
1	i-pr	Bu	c-Pr	2.6	0.41497
2	NMe_2	Bu	c-Pr	5.6	0.74818
3	NMe_2	Pr	c-Pr	12	1.07912
4[c]	NMe_2	Pr	Ph	3.6	0.55631
5	N(Me)COO i-Bu	Bu	c-Pr	7.4	0.86921
6	N(Me)COO i-Bu	Pr	Ph	0.5	−0.301
7	–	Bu	Ph	3	0.47712
8	–	Bu	(4-F)Ph	4	0.60205
9	–	Bu	2-Thieny1	8	0.90308
10	–	Bu	c-Pr	2.6	0.41497
11[c]	–	Bu	c-Pentyl ethyl	1.5	0.17602
12	–	Bu	(1-Me)Cyclo-Pr	1.1	0.04132
13[c]	–	Bu	$-(CH_2)_5NH$-Boc	12	1.07912
14	–	Bu	$-(CH_2)_5NH_2$	4.4	0.64341
15	–	Pr	Cyclo-Pr	2.8	0.44715
16[c]	–	Pr	$-CH_2)_4COOH$	25	1.39797
17	–	Bu	NH-i-Pr	5.6	0.74812
18	–	Bu	NH-Ph	5.9	0.7708
19[c]	–	Bu	Tetrazole	5	0.6989

[a] Inhibition of ^{125}I Angiotensin II binding to rabbit aorta membrane receptor (AT_1 receptors) (IC_{50}).
[b] pIC_{50} to generate equation.
[c] Compounds belonging to test set.

to QSAR. Most stable structure for each compound was generated after energy minimization and used for calculating various physico-chemical descriptors like thermodynamic, steric and electronic. The energy-minimized geometry was used for the calculation of the various 2D descriptors (Individual, Chi, ChiV, Path count, Chi Chain, ChiV Chain, Chain path count, Cluster, Path cluster, Kappa, Element Count, Estate number, Estate contribution, Semi-empirical, Hydrophillic–hydrophobic, Polar surface area and Alignment independent) and was considered as independent variables in the present study (Table 2). QSAR analysis was performed after removal of all the invariable columns, as they do not contribute to the QSAR. The optimal test and training data set were generated using the manual selection method. Sphere exclusion method (Golbraikh and Tropsha, 2002) was adopted for division of training and test set. Sphere exclusion method is used for creating training and test set from the data. This is a rational selection method which takes into consideration both biological and chemical space for division of dataset. Dissimilarity value provides handle to vary train/test set size. It needs to be adjusted by trial and error until a desired division of train and test set is achieved. As a rule, increase in dissimilarity value will lead to increase in number of molecules in the test set. A commonly used ratio of training to validation objects (test set), which was also adopted in this work, is 75:25%. All 19 molecules were subjected to regression analysis using partial least square analysis, as model building methods coupled with Genetic algorithms. Regression analysis was carried out for Angiotensin II receptor activity and the best model was cross-validated.

2.3. Three dimensional QSAR studies

In the kNN-MFA method, two models were generated for the selected members of training and test sets, and the corresponding best models are reported herein. VLife Molecular Design Suite 3.5 allows user to choose probe, grid size, and grid interval for the generation of descriptors. The variable selection methods along with the corresponding parameters are allowed to be chosen, and optimum models are generated by maximizing q^2. k-Nearest neighbor molecular field analysis (kNN-MFA) requires suitable alignment of given set of molecules. To derive the kNN-MFA descriptor fields, a 3D cubic lattice with grid spacing of 2 Å in x, y, and z dimensions was created to encompass the aligned molecules. kNN-MFA descriptors were calculated using an sp^3 carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 with default cut-off energy 30 kcal/mol to generate steric field, electrostatic

Table 2 Selected 2D QSAR of biphenyl acylsulfonamides derivatives.

SsssCHE index	SddCE-index	SsClE	Smr	Chi4pathCluster	T_2_Cl_4
1.086669	0.316538	0.749299	11.80996	97.321	25
1.226031	0.533764	0.896488	6.221686	110.52	25
1.209908	0.456622	0.878385	11.07407	119.31	25
1.087406	0.309316	0.747295	11.61199	113.28	25
1.126616	0.380664	0.784978	12.01505	112.28	25
1.016766	0.209136	0.687359	11.83269	113.07	25
1.196249	0.442516	0.854882	13.56656	107.28	25
1.09618	0.298983	0.744843	11.94554	93.135	25
1.180693	0.587311	0.888138	9.145759	102.65	27
1.101994	0.520555	0.801299	5.849149	116.28	25
1.254364	0.643208	0.925044	6.315244	101.89	23
1.13186	0.529668	0.840457	5.925926	97.23	25
1.20383	0.594758	0.902644	7.001888	87.05	25
1.00671	0.292575	0.690153	10.93499	113.07	25
1.150279	0.443934	0.758491	13.02988	96.28	25
1.196024	0.575352	0.847258	5.830085	90.29	25
1.163614	0.497534	0.839842	13.23062	130.75	25
1.112297	0.457762	0.748842	12.54838	99.08	25
1.107289	0.5012	0.800415	5.955034	72.15	25

and hydrophobic fields. The steric, electrostatic and hydrophobic energy values were truncated at a default value of ±30 kcal/mol. The kNN-MFA steric, electrostatic and hydrophobic fields thus generated were scaled by the standard method in the software. The 3D-QSAR studies were performed by kNN-MFA using Genetic algorithms selection method. Multiple conformation of each molecule was generated using the Monte Carlo conformation search method. It is a random search method for finding the conformations of molecules (Metropolis et al., 1953), which uses the metropolis condition to accept or discard generated conformers. The software produced more than 7568 descriptors and prior to model development descriptors having zero values or same values were removed which resulted in more than total 2500 descriptors for all the compounds in separate columns (Table 3). This

Table 3 Selected 3D QSAR descriptors of biphenyl acylsulfonamides derivatives.

S_590	S_628	H_1340	H_524	E_380
−0.00145	−0.00185	0.311043	0.288394	−0.34491
−0.0014	−0.00179	0.317634	0.288597	−0.73824
−0.00133	−0.00169	0.264997	0.242913	−0.18831
−0.00155	−0.00194	0.537642	0.497805	−0.06079
−0.00117	−0.00192	0.390055	0.357521	−0.20845
−0.00147	−0.00189	0.383007	0.351894	−0.04222
−0.00136	−0.00177	0.396775	0.366952	−0.10001
−0.00121	−0.00156	0.324362	0.300951	−0.32251
−0.00138	−0.00178	0.300821	0.278123	−0.15699
−0.00159	−0.0021	0.425603	0.396065	−0.22522
−0.00137	−0.00176	0.305052	0.281573	−0.21476
−0.00132	−0.00172	0.420478	0.391495	−0.15651
−0.00366	−0.00388	0.421436	0.388603	−0.27243
−0.00324	−0.00341	0.360514	0.331822	−0.45635
−0.00368	−0.00387	0.344948	0.311835	−0.19449
−0.00437	−0.0048	0.386581	0.366657	−0.00796
−0.00363	−0.00376	0.3448	0.310924	−0.35293
−0.00355	−0.00375	0.49126	0.450296	−0.35293
−0.015124	−0.00578	0.2987	0.343213	−0.32547

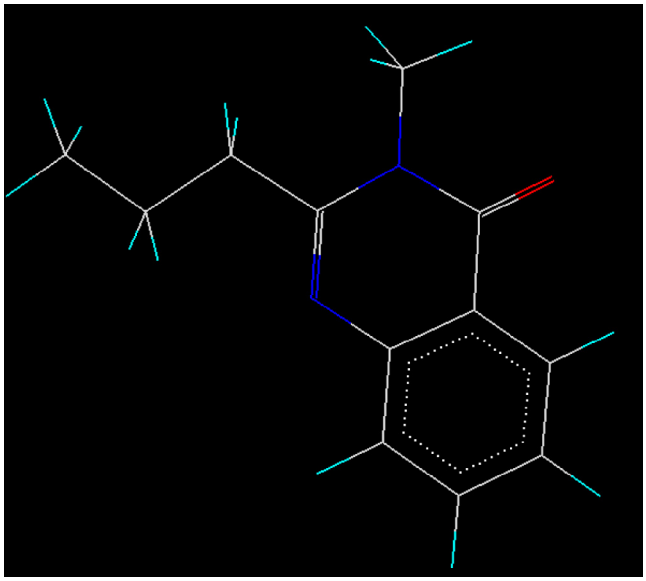


Figure 1a Template molecules.

algorithm allows constructing training sets covering all descriptor space areas occupied by representative points. kNN-MFA (Ajmani et al., 2006) with Genetic algorithms and stepwise variable selection was employed for selection of variables to obtain the QSAR model. The standard leave-one-out (LOO) procedure was implemented to calculate cross validated r^2 (q^2) value, that is a molecule in the training set was eliminated and its biological activity was predicted as the weighted average activity of the k most similar molecules.

2.3.1. Molecular alignment

Molecular alignment was used to visualize the structural diversity in the given set of molecules. This was followed by generation of common rectangular grid around the molecules. The template structure, i.e. quinazolinone ring was used for

alignment (Ajmani et al., 2006) by considering the common elements of the series as shown in Fig. 1a. The reference molecule 8 is chosen high antihypertensive activity which made it a valid lead molecule and therefore was chosen as a reference molecule. After optimizing, the template structure and the reference molecule were used to superimpose all molecules from the series using the template alignment method. kNN-MFA method requires suitable alignment of given set of molecules after optimization; alignment was carried out by template based alignment method (Fig. 1b).

2.3.2. k-Nearest neighbor method

The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures). This method employs the kNN classification principle combined with the Genetic algorithms selection procedure for optimization of the number of nearest neighbors (k) used to estimate the activity of each compound and optimization of selection of variable from the original pool of all molecular descriptors (steric and electrostatic fields at the lattice points) that are used to calculate similarities between compounds (Sharaf et al., 1986). The descriptors that get selected in a given model are the field points either of steric or of electrostatic nature at particular locations in a common grid around reported set of molecules. For utilizing these descriptors for new ligand design, we consider the field values at different grid points of compounds cluster having most active compound. The extrema of field values of compounds in the cluster of most active compounds decide the range of field values which is preferred and recommended for new compound design.

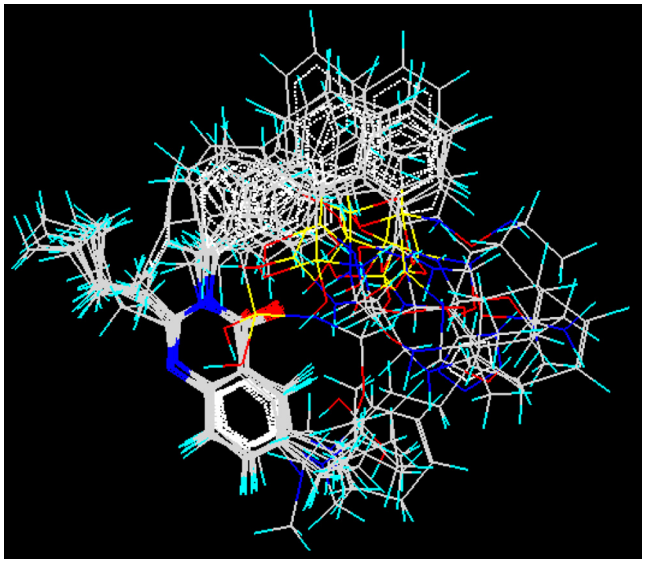


Figure 1b 3D view of aligned molecules.

2.3.3. kNN-MFA with Genetic algorithm

The genetic function approximation (GFA) algorithm offers a new approach to the problem of building quantitative structure–activity relationship (QSAR) and quantitative structure–property relationship (QSPR) models. Unlike most other analysis algorithms, GFA as a result provides multiple models, and the population of the model is created by evolving random initial model using a Genetic algorithm. Genetic algorithms (GA) described by Holland, is one of the most popular stochastic optimization techniques that mimic natural evolution and selection (Holland, 1992). It is a class of algorithms inspired by the process of natural evolution in which species having a high fitness under some conditions can prevail and survive to the next generation; the best species can be adapted by crossover and/or mutation in the search for better individuals. In this method, a chromosome and its fitness in the species represent a set of molecular descriptors and the cross-validated predictive accuracy of the derived QSAR model respectively. Here, a solution is typically encoded as a bit string consisting of a series of the digits 1 and 0. The GA begins by generating a set of random solutions (the population), which are analogous to a set of chromosomes in a biological system. The set of

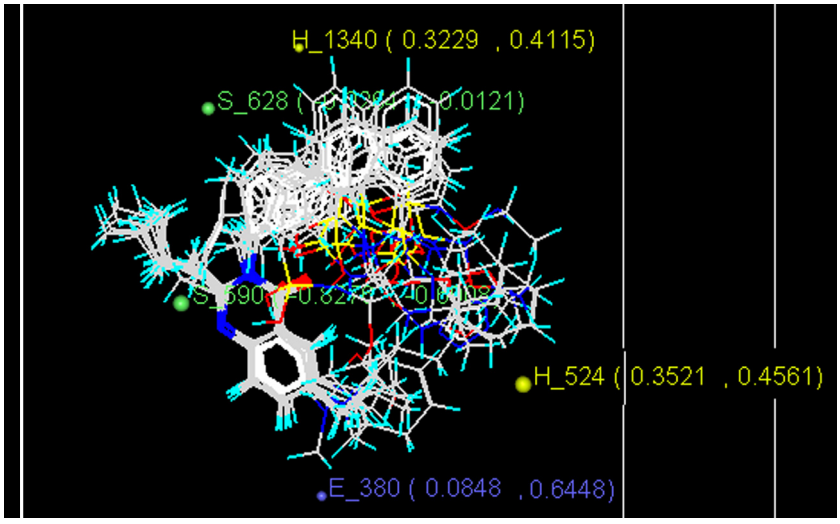


Figure 1c Contribution plot of 3D QSAR GA-kNN-MFA.

variables indicated with a value of 1 in the chromosome is then used as input for model building (e.g., PLS). The reproductive population and new population are combined and mutated with a predefined mutation rate. The best chromosome in the reproductive population is kept from the mutation process. This process is repeated over a number of 'generations' until the population converges to a small set of solutions. The cycle is repeated until the number of generations reaches a given maximum. The final model obtained is further refined by removing descriptors which do not affect predictive accuracy significantly. In this Genetic algorithms search for 3D QSAR, cross correlation limit was set as 0.5, population and number of generations as 10 and 1000, respectively and speed of 999 whereas in 3D QSAR, cross correlation limit was set as 0.5, population and number of generations as 100 and 500, respectively and speed of 999. Genetic algorithms are especially good at searching problem spaces having a large number of dimensions, since they conduct a very efficient, directed sampling of the large space of possibilities.

2.3.4. Stepwise (SW) forward variable selection

In stepwise (SW) forward variable selection algorithm, the search procedure begins with developing a trial model step by step with a single independent variable and to each step; independent variables are added one at a time, examining the fit of the model by using the PLS cross-validation procedure. Thus, the model is repeatedly altered from the previous one by adding or removing a predictor variable in accordance with the 'stepping criteria' (in this case $F = 4$ for inclusion; $F = 3.99$ for exclusion for the forward–backward selection method). The method continues until there is no more significant variable remaining outside the model.

2.4. Development and validation of QSAR models

Models I, II, and III were generated by using significant statistical methods, namely, partial least square (PLS). PLS is an effective technique for finding the relationship between the properties of a molecule and its structure. In mathematical terms, PLS relates a matrix Y of dependent variables to a matrix X of molecular structure descriptors, i.e., a latent variable approach to modeling the covariance structures in these two spaces. PLS have two objectives: to approximate the X and Y data matrices, and to maximize the correlation between them. Whereas the extraction of PLS components is performed stepwise and the importance of a single component is assessed independently, a regression equation relating each Y variable with the X matrix is created. PLS decomposes the matrix X into several latent variables that correlate best with the activity of the molecules.

The cross-validation analysis was performed using the leave-one-out method. The following statistical parameters were considered to compare the generated QSAR models: correlation coefficient (r), squared correlation coefficient (r^2), predicted r^2 (pred_r^2), and Fischer's value (F). To validate the generated QSAR models, the leave-one-out (LOO) method was used, indicated as the value of q^2 (cross-validated explained variance), (Cramer et al., 1988) which is a measure of the internal predictive ability of the model. The cross-validated r^2 (q^2) value was calculated, where y_i and $\hat{y}_i$ are the actual and predicted activities of the ith molecule, respectively, and y_{mean} is the average activity of all the molecules in the training set. Both summations are over all molecules in the training set and hence the predictions were based on the current trial solution, the q^2 obtained indicates the predictive power of the current model.

$$q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{\text{mean}})^2}$$

The predicted r^2 (pred_r^2) value was calculated, where y_i and $\hat{y}_i$ are the actual and predicted activities of the ith molecule in test set, respectively, and y mean is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. The pred_r^2 value is indicative of the predictive power of the current model for external test set.

$$\text{pred\_}r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{\text{mean}})^2}$$

To evaluate the statistical significance of the QSAR model for an actual data set, we have employed a one-tail hypothesis testing. The robustness of the QSAR models for experimental training sets was examined by comparing these models to those derived for random data sets. Random sets were generated by rearranging biological activities of the training set molecules.

3. Results and discussion

The importance and utility of the new 2D and 3D QSAR method discussed has been established by applying it to known sets of molecules as described above. All the calculated descriptors were considered as independent variable and biological activity as dependent variable. In 2D QSAR analysis, significant methods partial least square analysis, were applied to generate three models: Models I, II, and III, respectively, from these models, three of them were having good q^2 and pred_r^2 values, one of which was selected having good internal and external predictivity. Selecting training and test set was by spherical exclusion method. The QSAR models developed by kNN-MFA include both the electrostatic, steric and hydrophobic descriptors along with their range to indicate their importance for interaction in molecular field. Models IV and V 3D QSAR Genetic algorithms studies. QSAR investigations of the substituted quinazolinone biphenyl acylsulfonamides derivatives series resulted in several QSAR equations. Some statistically significant 2D and 3D QSAR models were chosen for discussion.

3.1. Model I

$pIC_{50} = +2.76343(\pm 0.5418)$ SsssCHE index $-0.87432(\pm 0.1126)$ SddCE-index $+0.6421(\pm 0.0435)$ T_2_Cl_4 $+0.9751(\pm 0.0431)$ SssNHE-index $+4.129$.

$N = 19$, Degree of freedom $= 20$, $r^2 = 0.8056$, $q^2 = 0.6742$, F test $= 71.210$, r^2se $= 0.4398$, q^2se $= 0.2532$, pred_$r^2 = 0.7583$, and pred_r^2se $= 0.2165$.

The derived model shows good correlation between biological activity and parameters SsssCHE index, SddCE-index, T_2_Cl_4, and SssNHE-index as the coefficient of determination (r^2) = 0.8056 capable of explaining 80% of variance in the observed activity values. The low standard error of r^2_se = 0.4398 demonstrates accuracy of the model. The

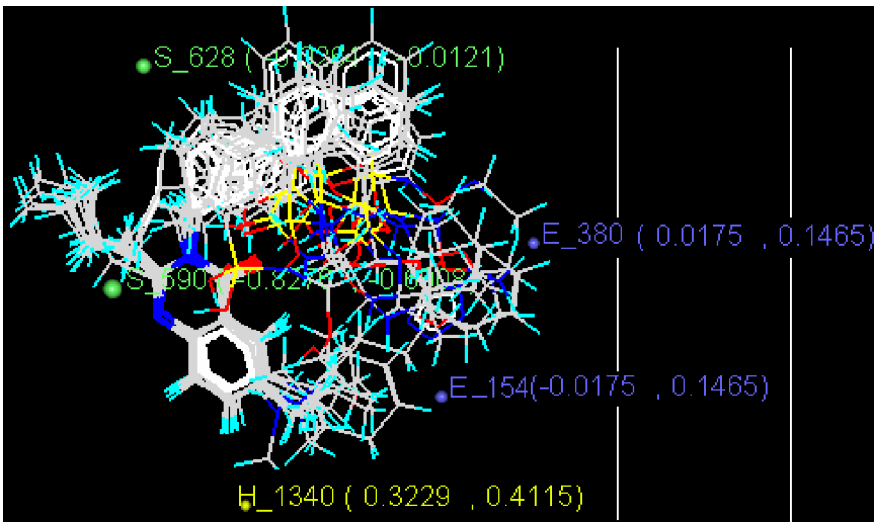


Figure 1d Contribution plot of 3D QSAR SW-kNN-MFA.

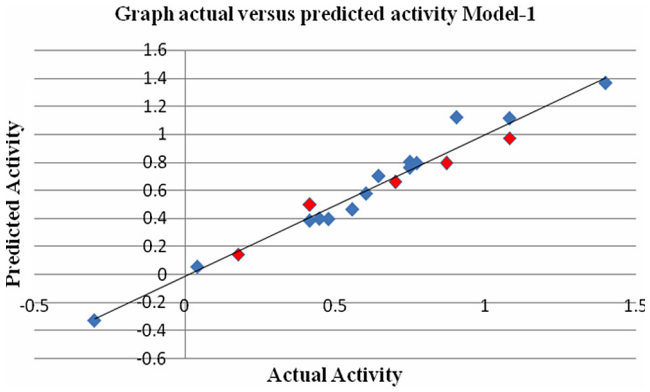


Figure 2a Graph of actual activity versus predicted activity Model I.

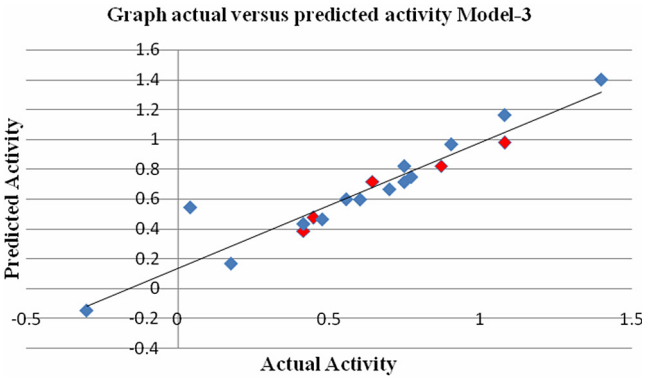


Figure 2c Graph of actual activity versus predicted activity Model III.

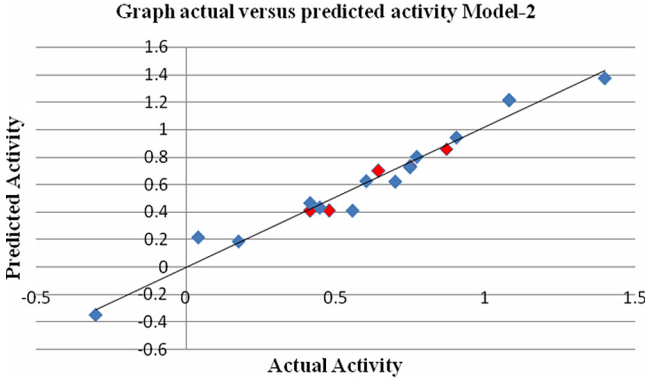


Figure 2b Graph of actual activity versus predicted activity Model II.

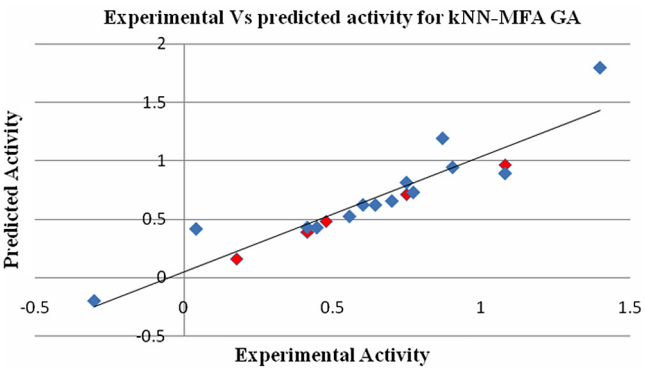


Figure 2d Graph for experimental versus predicted activity for kNN-MFA_GA.

leave-one-out procedure was used for internal validation of the model. The model showed an internal predictive power cross validated r^2 (q^2 = 0.6742) of 67.42% and predictivity for external test set (pred_r^2 = 0.7583) about 75.83% and low q^2_se = 0.2532 values reflect good internal predictive power of the model. The *F*-test = 71.210 shows the overall statistical significance level of 99.99% of the model which means the probability of failure of the model is 1 in 10,000. The graph of actual versus predicted activity is shown in Fig. 2a. Since highest positively contributed descriptor in the above model SsssCHE index

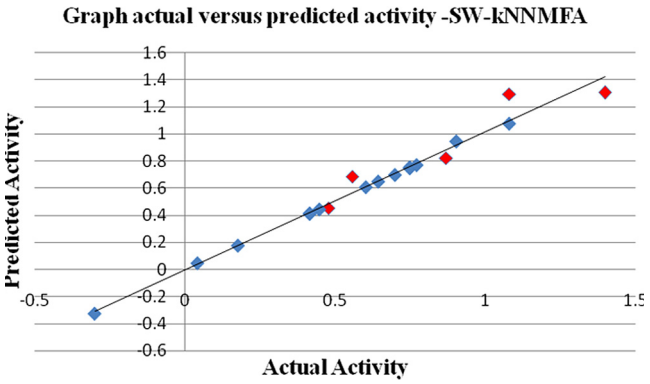


Figure 2e Graph for experimental versus predicted activity for kNN-MFA_SW.

signifies total number for number of —CH group connected with three single bonds. The positive correlation suggests that Angiotensin II activity quinazolinone biphenyl acylsulfonamides derivatives may be increased by increasing the number of such —CH groups present in the molecules. Further it may be inferred that increasing the saturation of the aromatic rings will be contribute to increases in Ang II activity. SddCE-index defines electrotopological state indices for number of carbon atom connected with two double bonds. The descriptors show negative correlation among the parameters selected for the derived QSAR model. The negative coefficients suggest that inclusion of such carbon atoms in the molecules lead to increased to decreased Ang II activity. SssNHE-index describes electro topological state indices for number of —NH group connected with two single bonds. Its positive contribution in the QSAR model implies that will lead to increases potency for the instead of COOH group. Its positive value suggests that increasing the number of such carbons will lead to better Ang II potency. This type of electro topological property provides flexibility hence better fitting into the receptor cavity. The descriptor T_2_Cl_4 is the number of double-bonded atoms separated from the chlorine atom by four bonds. It is another influential alignment-independent descriptor suggesting that the presence of substituents with chlorine on the phenyl ring at the ortho position will lead to an increase in activity. The inter-correlation among the selected descriptors was very less due to auto scaling and cross correlation limit permitted was 0.6. The correlation matrix is shown in Table 5, which shows good correlation of selected parameters with biological activity. The above model is validated by predicting the biological activities of the test molecules, as indicated in Table 4.

Table 5 Correlation matrix for the QSAR model (Model I).

	SsssCHE index	SddCE-index	T_2_Cl_4	SssNHE-index
SsssCHE index	1.0000			
SddCE-index	0.35741	1.0000		
T_2_Cl_4	0.30613	0.60543	1.0000	
SssNHE-index	0.16865	0.49832	0.72187	1.0000

3.2. Model II

pIC_{50} = +0.3551(±0.0748) Chi4pathCluster +0.1642(±0.0470) $SssNH_2E$-index +0.2489(±0.0195) SsClE-index +0.0445.

N = 19, Degree of freedom = 19, r^2 = 0.7329, q^2 = 0.6912, F test = 56.019, r^2se = 0.4336, q^2se = 0.5476, pred_r^2 = 0.7185, and pred_r^2se = 0.6598.

To improve the external predictivity of the model, PLS analysis with the data set was performed, Model II is a triparametric model generated with coefficient of determination (r^2) = 0.7329 which is capable of explaining 73.29% of variance in the observed activity values. The model showed an internal predictive power (q^2 = 0.6912) of 69.12% and predictivity for external test set (pred_r^2 = 0.7185) about 71%. The F-test = 56.019 shows the overall statistical significance level of 99.99% of the model which means the probability of failure of the model is 1 in 10, 000. In accordance with PLS model the descriptor Chi4pathCluster, this descriptor signifies molecular

Table 4 2D QSAR and 3D QSAR observed and predicted (LOO) activity of derivatives.

Observed activity	Predict model 2D QSAR-I	Predict model-2D QSAR-II	Predict model-2D QSAR-III	Predict model-GA-kNN	Predict model-SW-kNN
0.4149	0.4968	0.4115	0.3032	0.4495	0.4086
0.7481	0.8032	0.7324	0.8214	0.7129	0.7454
1.0791	0.9732	0.9815	0.9786	1.1341	1.2961
0.5563	0.4654	0.4104	0.5983	0.5236	0.6854
0.8692	0.7938	0.8574	0.8214	0.9125	0.8216
0.3010	0.3255	0.1497	0.2958	0.2693	0.3286
0.4771	0.3968	0.4118	0.4632	0.4802	0.4531
0.6020	0.5783	0.6254	0.5968	0.6126	0.6053
0.9030	1.121	0.9432	0.9683	0.9066	0.9432
0.4149	0.3857	0.4654	0.4327	0.4189	0.4115
0.1760	0.1432	0.1857	0.1659	0.1721	0.1741
0.0413	0.0564	0.2154	0.5439	0.2165	0.0438
1.0791	1.1136	1.2164	1.1652	1.0736	1.0743
0.6434	0.7032	0.6994	0.7165	0.6395	0.6468
0.4471	0.3998	0.4327	0.4738	0.4277	0.4408
1.3979	1.3654	1.3753	1.4032	1.3914	1.3034
0.7481	0.7632	0.7265	0.7143	0.7439	0.7513
0.7708	0.7948	0.8021	0.7482	0.7987	0.7682

connectivity index of 4th-order pathcluster and encodes structure information specifically on a branch point, emphasizing the immediate branch point environment. It is defined for the branched skeleton and positive contribution of this descriptor in the QSAR model demonstrates that increased branching will lead to improved Ang II activity. The SsClE-index is an electro-topological parameter which can define the total number of chlorine atoms connected with one single bond is which also contributing immensely to activity. The positive coefficient of the descriptor suggests that Ang II activity of quinazolinone derivatives may be increased by increasing the number of chlorine atoms present in the nucleus. $SssNH_2E$-index describes electro topological state indices for number of $-NH_2$ group connected with two single bonds. Its positive value suggests that increasing the number of such NH group will lead to better Ang II potency. The correlation plot of the actual versus predicted activity is shown in Fig. 2b. The cross-correlation between the pivotal descriptors is presented in Table 6. The above model is validated by predicting the biological activities of the test molecules, as indicated in Table 4.

3.3. Model III

$pIC_{50} = +0.1293(\pm 0.0370)$ smr $-0.1646(\pm 0.0476)$ 4PathCount $+0.1030(\pm 0.0524)$ Rotatable Bond Count $+0.3116$.

$N = 19$, Degree of freedom = 18, $r^2 = 0.7028$, $q^2 = 0.6361$, F test = 26.528, r^2se = 0.1421, q^2se = 0.8094, pred_r^2 = 0.6814, and pred_r^2se = 1.7431.

The derived model shows good correlation between biological activity and parameters smr, 4PathCount and Rotatable Bond Count as the correlation coefficient $r^2 = 0.7028$ and the model explains about 70% variance in Ang II activity exhibited by quinazolinone derivatives. The low standard error of r^2_se = 0.1421 demonstrates accuracy of the model. The model shows overall significance level better than 99.99%, with $F = 26.528$ against values of 99.99% significance. The leave-one-out procedure was used for internal validation of the model. In this procedure high cross validated r^2 ($q^2 = 0.6361$) and low q^2_se = 0.8094 value, reflects the very good internal predictive power of the model. However, a high q^2 value does not necessarily give a suitable representation of the real predictive power of the model for Ang II inhibitory ligands. So, an external validation was also carried out in the present study. Another parameter for predictivity of test set compound is high pred_r^2 = 0.6814 and low pred_r^2se = 1.7431, which is showing good external predictive power of the model. The results are listed in 4 and the correlation plot of the actual versus predicted activity is shown in Fig. 2c. Smr descriptor signifies the molecular refractivity (including implicit hydrogens). This property is an atomic contribution model that assumes the correct protonation state (washed structures). It's positive contribution suggests that increment of the polarity of the molecule leads to increased Ang II activity. 4PathCount is a topological parameter which can signify the total number of fragments of fourth order (four bondpath) in compound. It is negatively correlated with inhibitory activity so, it may be inferred that increasing the branching of compound is favorable for Ang II activity. The Rotatable Bond Count refers to number of rotatable bonds in the molecules. The slightly positive term associated with the descriptor in QSAR model indicates that fractional increase in the rotatable bonds in the molecule is beneficial for Angiotensin II activity exhibited by all the compounds. The above model is validated by predicting the biological activities of the test molecules, as indicated in Table 4. The cross-correlation between the pivotal descriptors is presented in Table 7.

3.4. Model IV-3D QSAR Genetic algorithms

$pIC_{50} = -1.4364 - (-0.0264, -0.0121)$ S_628 + (0.3229, 0.4115) H_1340 − (−0.8276, −0.6008) S_590 + E_380 (0.0848, 0.2019) + (0.3521, 0.4561) H_524 + 3.9281.

$N = 19$, Degree of freedom = 18, $q^2 = 0.7263$, F test = 35.632, r^2se = 0.7643, q^2se = 0.5312, pred_r^2 = 0.6693, and pred_r^2se = 1.29876.

The descriptors that get selected in a given model are the field points either of steric, electrostatic, and hydrophobic nature at particular locations in a common grid around reported set of molecules. For 3D QSAR a kNN-MFA of substituted quinazolinone biphenyl acylsulfonamides derivatives with reported activities against the Ang II receptor antagonists was performed. Genetic algorithms method resulted in several statistically significant models, of which the corresponding best model, IV is reported herein. The model selection criterion is the value of q^2, the internal predictive ability of the model, and that of pred_r^2, the ability of the model to predict the activity of external test set. As the cross-validated correlation coefficient (q^2) is used as a measure of reliability of prediction, the correlation coefficient suggests that our model is reliable and accurate. A data set of compounds containing five molecules was selected as the test set from the original data of 19 compounds for the validation experiments. Since the calculation of the pairwise molecular similarities and hence the prediction was based upon current training set, the q^2 value obtained (0.7263) is the indicative power of the current kNN-MFA model. The above steps were repeated for $k = 4$, etc. (upper limit of k is the total number of molecules in the data set). From these models, the one with $q^2 \geqslant 0.5$ and pred_$r^2 \geqslant 0.5$ was selected to predict the activity of training and test set of compounds. According to the kNN methodology pIC_{50} is a function of independent variables, steric, electrostatic and hydrophobic fields. Values of k (4), q^2 (0.7263), pred_r^2 (0.6693), q^2_se (0.5312), and pred_r^2se (1.29876) prove that QSAR equation so obtained is statistically significant and shows the predictive power of the model is 72.63% (internal

Table 6 Correlation matrix for the QSAR model (Model II).

	Chi4pathCluster	$SssNH_2E$-index	SsClE-index
Chi4pathCluster	1.0000		
$SssNH_2E$-index	0.42944	1.0000	
SsClE-index	0.53278	0.7754	1.0000

Table 7 Correlation matrix for influencing the biologic activity (Model III).

	Smr	4PathCount	Rotatable bond
Smr	1.0000		
4PathCount	0.23824	1.0000	
Rotatable bond	0.58633	0.78041	1.0000

validation). The negative range and negative value of steric descriptors as S_628 (−0.0264, −0.0121), S_590 (−0.8276, −0.6008) chosen by sphere exclusion method of data selection and forward backward method of variable selection indicating its crucial role in predicting Ang II activity and signifying negative range of steric descriptors indicate that negative steric potential is favorable for activity and less bulky substituent is preferred in that region. The graphical representations and model summary of kNN-MFA results for Ang II activity are shown in Fig. 1c. The plot of the kNN-MFA shows the relative position and ranges of the corresponding important electrostatic, steric and hydrophobic fields in the model which provide guidelines for new molecule design (Table 4). The steric effect, as shown with green color around the quinazolinone ring, implies about the preferred substitution (less or more bulky group) to produce higher Ang II activity. The increase in negative value of S_590 shows that quinazolinone ring (R_1) is important for activity and it can be replaced with less bulky ring substituent. E_380 (0.0848, 0.2019) Electrostatic descriptor with positive coefficient (E_380) around position of the quinazolinone ring corroborates that electropositive (electron-withdrawing) group is preferred at R_2-position of quinazolinone ring. Positive contribution of H_1340 and H_524 to hydrogen group nearer to R_1 and R_3, respectively indicates that positive hydrophobic field is favorable for increasing the activity. Hence less hydrophobic or more hydrophilic substituent groups near R_1 and R_3 are preferred. The GA-kNN model field plot and corresponding important steric fields range which show the ranges are more toward the negative side meaning decreasing steric bulk of the substituent group is favorable at the respective substitution site. The correlation plot of the actual versus predicted activity is shown in Fig. 2d.

3.5. Model V-3D QSAR SW-kNN-MFA

pIC_{50} = E_154 (0.0175 0.1465) − E_380 (0.0848 0.2019) + H_1340 (0.3229, 0.4115) − S_628 (−0.0264–0.0121) − S_590 (−0.8276–0.6008).

N = 19, Optimum Components = 3, DF = 20, r^2 = 0.8316, q^2 = 0.6516, F test = 32.895, r^2_se = 0.3176, q^2_se = 0.6954, pred_r^2 = 0.6954, pred_r^2se = 0.5387, Zscore $\bar{Q}^2$ = 1.321, Best Rand $\hat{Q}^2$ = 0.7643.

The descriptors E_154, E_380, H_1340, S_628, and S_590 are the steric and electrostatic field energy of interactions between probe (CH_3) and compounds at their corresponding spatial grid points of 154, 380, 1340, 628, and 590. The respective relative contribution of steric, electrostatic, and hydrophobic fields indicates that electrostatic field is more predominant and contribution chart of selected descriptors are represented in Fig. 1d. It is evident that the predicted activities of all the compounds in the test set are in good agreement with their corresponding experimental activities and optimal fit is obtained (Table 4). The contribution plot of steric, electrostatic and hydrophobic field interactions indicates relative regions of the local fields (steric, electrostatic and hydrophobic) around the aligned molecules, leading to activity variation in the model. The green-colored balls specify the positions of the steric descriptors and the descriptors with positive or negative coefficients show a region where bulky substituent is favored or disfavored, respectively. Electrostatic field descriptors (blue-colored balls) with positive coefficients represent regions where electropositive (electron-withdrawing) groups are favorable, whereas negative coefficient indicates that electronegative (electron-rich or electron-donating) groups are favorable in this region (Samee et al., 2008). From 3D-QSAR model it is observed that electrostatic descriptors like E_154 (−25.04%) and E_380 (−19.42%) with negative coefficient are from the R_1 and R_2 position of the quinazolinone biphenyl acylsulfonamides ring. H_1340 descriptors to Hydrogen group nearer to R_1 and R_2 respectively indicates that positive hydrophobic field is favorable for increasing the activity. Hence less hydrophobic or more hydrophilic substituent groups near R_1 and R_2 are preferred. The correlation plot of the actual versus predicted activity is shown in Fig. 2e. The robustness of the QSAR model for experimental training sets was examined by comparing this model to those derived for random dataset. The QSAR model was evaluated using the following statistical measures; numbers of observations, i.e., molecules in data set (n = 19); numbers of nearest neighbors (k = 3); cross validated r^2 (q^2 = 0.6516); predicted r^2 for the external test set (pred_r^2 = 0.6954); standard error of r^2 (q^2_se = 0.6954); standard error of cross validated r^2 (pred_r^2se = 0.5387).

4. Conclusions

In the present investigation, all proposed QSAR models were statistically significant, thus, from above QSAR investigations it could be concluded that 2D/3D descriptors properties of substituted quinazolinone biphenyl acylsulfonamides derivatives are mainly involved in eliciting Angiotensin II receptor activity. The good correlation between experimental and predicted biological activity for five compounds in the test set further highlights the reliability of the constructed QSAR model. The requirements for the anti-hypertensive activity are explored with 2D, 3D and group based QSAR studies. The 2D technique indicates the importance of SssNHE-index, T_2_Cl_4, $SssNH_2$E-index, SsClE-index and Rotatable Bond Count of the compounds on the activity. The 3D QSAR analysis makes it possible to relate chemical structures of ligands and their binding affinity with respect to different bio targets by using the kNN-MFA techniques. Thus it provides a direct view of factors expressed in terms of molecular fields (electrostatic, steric, and hydrophobic) affecting the binding affinity. This in turn could give the reasonably good prediction of binding affinity. The location and range of function values at the field points selected by the models provide clues for the design of new molecules. Hence, this method is expected to provide a good alternative for the drug design.

Acknowledgment

The author wishes to express gratitude to V-life Science Technologies Pvt. Ltd. for providing the software for the study.

References

Ajmani, S., Jhadav, K., Kulkarni, S.A., 2006. J. Chem. Inf. Model. 46, 24–31.

Belvisi, L., Bravi, G., Catalano, G., Mabiliab, M., Salimbeni, A., Scolastico, C., 1996. J. Comput. Aided Mol. Des. 10, 567–582.

Burnier, M., Brunner, H.R., 1997. Expert Opin. Investig. Drugs 6, 489–500.
Chakravarty, P.K., Strelitz, R.A., Chen, T.Bau., Chang, R.S.L., Victor, J.L., Zingaro, G.J., Schom, T.W., Kivlighn, S.D., Siegl, P.K.S., Patchettt, A.A., Greenleet, W.J., 1994. Bioor. Med. Chem. Lett. 4 (1), 75–80.
Cramer, R.D., Patterson, D.E., Bunce, J.D., 1988. J. Am. Chem. Soc. 110, 5959–5967.
Datar, P.A., Desai, P.V., Coutinho, E.C., 2004. J. Chem. Inf. Model. 44, 210–220.
Dlaz, H.G., Prado, F.J., 2008. J. Comput. Chem. 29, 656–667.
Dzau, V.J., Pratt, R.E., 1986. Renin-angiotensin system: biology, physiology and pharmacology. In: Fozzard, H.A., Haber, E., Jennings, R.B., Katz, A.M., Morgan, H.E. (Eds.), The Heart and Cardiovascular System. Raven Press, New York, pp. 1631–1662.
Ellingboe, J.W., Antane, M., Nguyen, T.T., Collini, M.D., Antane, S., Bender, R., Hartupee, D., White, V., McCallum, J., Park, C.H., 1994. J. Med. Chem. 37, 542–550.
Ferrario, C.M., 1990. J. Cardiovasc. Pharmacol. 15 (3), 51–55.
Gasteiger, J., Marsili, M., 1980. Tetrahedron 36, 3219–3228.
Golbraikh, A., Tropsha, A., 2002. J. Comput. Aided Mol. Des. 16, 357–369.
Halgren, T.A., 1996. J. Comput. Chem. 17, 553–586.
Holland, J.H., 1992. Sci. Am. 267, 66–72.
Ito, M., 2005. Nippon. Rinsho. 63 (3), 22–27.
Klebe, G., Abraham, U., Mietzner, T., 1994. J. Med. Chem. 37, 4130–4146.
Kurup, A., Garg, R., Carini, D.J., Hansch, C., 2001. Chem. Rev. 101, 2727.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. J. Chem. Phys. 21, 1087–1092.
Samee, W., Nunthanavanit, P., Ungwitayatorn, J., 2008. Int. J. Mol. Sci. 9, 235–246.
Sharaf, M.A., Illman, D.L., Kowalski, B.R., 1986. Chemometrics. Wiley, New York.
Sharma, M.C., Kohli, D.V., 2011a. World Appl. Sci. J. 12 (11), 2111–2119.
Sharma, M.C., Kohli, D.V., 2011b. World Appl. Sci. J. 12 (11), 2129–2134.
Sharma, M.C., Kohli, D.V., Chaturvedi, S.C., Sharma, S., 2009. Digest. J. Nanomat. Biostruct. 4 (4), 843–856.
Sharma, M.C., Sharma, S., Sahu, N.K., Kohli, D.V., 2011. Jour. Saud. Chem. Soc. 52.
Tuccinardi, T., Calderone, V., Rapposellim, S., Martinelli, A., 2006. J. Med. Chem. 49, 4305–4316.
VLife MDS software package, version 3.5, Pune, India 411007.
Yan, P.Z., Chen, B.C., Shi, H., Jian, H.J., Hai, L.W., Guo, L.S., Ru, Q.Y., 2007. Anal. Chim. Acta 593 (1), 68–74.
Yanagisawa, H., Amemiya, Y., Kanazaki, T., Shimoji, Y., Fujimoto, K., Kitahara, Y., et al., 1996. J. Med. Chem. 39, 323–338.